THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Exploring substitutability through discourse adverbials and multiple judgments

Citation for published version:
Rohde, H, Dickinson, A, Schneider, N, Louis, A & Webber, B 2018, Exploring substitutability through discourse adverbials and multiple judgments. in *IWCS 2017 - 12th International Conference on Computational Semantics.* ACL Anthology. <http://aclweb.org/anthology/W17-6814>

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
IWCS 2017 - 12th International Conference on Computational Semantics

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 07. May. 2023

Exploring Substitutability through Discourse Adverbials and Multiple Judgments

Hannah Rohde
University of Edinburgh
Hannah.Rohde@ed.ac.uk

Anna Dickinson
University of Edinburgh
Anna.Y.Dickinson@gmail.com

Nathan Schneider
Georgetown University
nathan.schneider@georgetown.edu

Annie Louis
University of Edinburgh
alouis@inf.ed.ac.uk

Bonnie Webber
University of Edinburgh
Bonnie.Webber@ed.ac.uk

Abstract

In his systematic analysis of discourse connectives, Knott (1996) introduced the notion of *substitutability* and the conditions under which one connective (e.g., *when*) can substitute for another (e.g., *if*) to express the same meaning. Knott only uses examples which he constructed and judged himself. This paper describes a new multi-judgment study on naturally occurring passages, on which substitutability claims can be tested. While some of our findings support Knott's claims, other pairs of connectives that Knott predicts to be *exclusive* are in fact judged to substitute felicitously for one another. These findings show that discourse adverbials in the immediate context play a role in connective choice.

1 Introduction

The question of how different discourse connectives are used to realize particular types of coherence relations remains unresolved. While some connectives show nearly one-to-one mappings with individual coherence relations, other connectives permit much more flexible usage across contexts.

One early enterprise targeting the above question was Alistair Knott's systematic assessment of the conditions that permit one connective to *substitute* for another (Knott, 1996). *Substitutability*, along with categories of coherence relations, then predicts the behavior of individual connectives. Another such enterprise is our own (Rohde et al., 2015, 2016, 2017) on implicit connectives in the context of explicit discourse adverbials. Using naturally occurring passages, we have gathered judgments from multiple participants as to what connective, if any, they could insert into a particular passage immediately before an existing discourse adverbial, to make explicit the author's intended message. For example, when shown the passage *It's too far to walk. Instead let's take the bus.*, a participant might insert *so* to express what she takes to be the intended causal reading.

Our findings show variation across participant responses. Such divergence in judgments could pose a puzzle for Knott's substitutability claims – Do they reflect (1) merely different interpretations of the passage or (2) genuine substitutability between the connectives selected? This paper reports a new study to address this puzzle. In the study, participants are asked to identify not only the connective that best expresses the intended meaning of a passage, but also *what other connectives* they could use to express the same meaning. The study makes three contributions: (1) It sheds light on our earlier data on divergences in participants' judgments; (2) it serves as a large-scale test of some of Knott's *substitutability* claims; and (3) it provides more evidence *against* the common assumption that an explicit discourse connective

between two clauses marks the coherence relation that holds between them and that no additional pragmatic inference is used to establish coherence between them (cf. Section 5). Correcting this assumption can improve modelling in computational semantics and the technology that depends on it, provide a more realistic account of translational divergences, and enable more effective design and interpretation of psycholinguistic experiments. As far as we are aware, the work presented here is the first to examine Knott's claims on the basis of large-scale experiments on naturally occurring data.

2 Background

2.1 Exploring discourse connectives through substitutability

In an innovative PhD thesis at the University of Edinburgh, Alistair Knott (1996) investigated what could be learned about discourse connectives through their possible substitutability relations. Informally, *substitutability* specifies the circumstances in which an author would be prepared to substitute one cue phrase for another in a passage of text (possibly with some reorganization of the passage and/or a change of style). So two cue phrases x and y may be *always* substitutable if wherever y appears, x is substitutable for y; *sometimes* substitutable if x is substitutable for y in some contexts in which y appears, but not all of them; or *never* substitutable if wherever y appears, x is not substitutable for y. His methodology involved: (1) Gathering a set of discourse connectives (which Knott called *cue phrases*); (2) defining a small set of *substitutability* relations which correspond to the contexts in which one cue phrase can substitute for another, with the same meaning being conveyed; (3) establishing the particular substitutability relations that hold between pairs of cue phrases; (4) using substitutability relations to define taxonomies of cue phrases; and (5) positing a set of semantic features that can be said to be intrinsic to cue phrases, such that subsumption relations between these features can explain the data-driven taxonomy of connectives based on substitutability.

The following three basic substitutability relations allow Knott to define four composite relations between cue phrases x and y, that underpin the rest of the thesis:

- SYNONYMOUS(x,y) if *always*(x,y) and *always*(y,x);
- EXCLUSIVE(x,y) if *never*(x,y) and *never*(y,x);
- HYPONYM(x,y) if *sometimes*(x,y) and *always*(y,x);
- CONTINGENTLY-SUBSTITUTABLE(x,y) if *sometimes*(x,y) and *sometimes*(y,x).

For example, SYNONYMOUS(*to begin with*, *to start with*) holds because in every context in which *to begin with* can be used as a cue phrase, so can *to start with*, with the same meaning being conveyed. In contrast, Knott claims EXCLUSIVE(*first*, *for one thing*) because there are no contexts in which *first* can substitute for *for one thing* and no contexts in which *for one thing* can substitute for *first*.

For the HYPONYM relation, Knott claims that HYPONYM(*for one thing*, *firstly*) holds because one can use *firstly* to start a sequence in any context, while one can only use *for one thing* to start a sequence in an argumentative context. Finally, the CONTINGENTLY-SUBSTITUTABLE relation is illustrated in (Knott and Mellish, 1996, p. 147) with *and* and *but* because there are some contexts in which *and* and *but* can both be used (Ex. 1), some contexts in which *and* can be used, but not *but* (Ex. 2), and some contexts in which *but* can be used, but not *and* (Ex. 3).

(1) Bill's a liar. He said he can run a mile in three minutes, [and, but] that's impossible.

(2) I'm very tired, [and, #but] I don't want to be disturbed.

(3) Don't be too harsh on Bob. He arrived late, [#and, but] he's usually very punctual.

Knott's analysis of cue phrases makes a further division of sense relations into ten sense *categories*: SEQUENCE, CAUSE, RESULT, RESTATEMENT, TEMPORAL, HYPOTHETICAL, SIMILARITY, DIGRESSION, ADDITIONAL INFORMATION and NEGATIVE POLARITY. Knott assigns some cue phrases to a

single category (when their only sense belongs to that category) and other cue phrases (e.g., *since*, *and*, *or*) to multiple categories (when they can be used to express more than one sense). Since *substitutability* of connectives within the same sentence requires the sentence to retain the same meaning, we assume that when two cue phrases do not share a single category in common, Knott would take them to be EXCLUSIVE. (While Spooren (1997) has posited specificity relations between sense categories, such that a CAUSAL relation can sometimes be conveyed by a TEMPORAL connective, our analysis here is based solely on Knott's empirical analysis involving *substitutability*.)

Knott's thesis incorporates ≈150 cue phrases into *substitutability* diagrams, which is a significant achievement. However, all examples in the thesis were ones he constructed, and all judgments, ones made by him alone. Knott recognized the need to carry out large-scale experiments using naturally occurring data to support his conclusions, but lacked the opportunity to do so.

Finally, as noted in point 5 above, while Knott posited a theoretical basis for substitutability in a set of binary-valued features intrinsic to the meaning of a cue phrase (or the sense of a cue phrase, for ones that belong to multiple categories), this paper just refers to Knott's claims about substitutability based on empirical judgments, and not to his later theoretical basis for the claims.

2.2 Collecting multiple judgments on discourse connectives

Our larger project addresses the common, but incorrect, assumption that only when explicit discourse connectives are absent or ambiguous is inference used to establish coherence between sentences and/or clauses. We have collected multiple judgments on connectives in naturally occurring text in order to understand and characterize implicit coherence relations that hold at the same time as coherence relations associated with explicit discourse adverbials. For example, while Ex. 4a contains only the explicit adverbial *instead*, it conveys the same meaning as Ex. 4b, in which the inferred causal relation has been made explicit.

(4) a. It's too far to walk. Instead let's take the bus.

b. It's too far to walk. So instead let's take the bus.

Both versions convey that we should take the bus as an alternative to walking because it's too far to walk.

Because judgments on discourse connectives can vary in unexpected ways, we have collected data on a large number of adverbials in a large number of passages from a large number of participants – one experiment using 20 discourse adverbials, aimed at understanding the extent of variability across adverbials, and another using 37 discourse adverbials, aimed at exploring adverbials in terms of their common paired-connective distribution.

2.2.1 Dataset of connective insertions

Our first study (Rohde et al., 2016) involved 28 naive participants, all native English speakers engaged through Amazon Mechanical Turk. We showed them passages with discourse adverbials and asked them to identify which of several given connectives (if any) could appear in the position before the adverbial, to explicitly signal their interpretation of the passage.

The target passages shown to participants (minimally, a sentence and maximally, a short paragraph) were selected from the *New York Times Annotated Corpus* (Sandhaus, 2008). Each target passage consisted of two spans of text, the second beginning with a discourse adverbial. Half the passages (*explicit passages*) originally contained a conjunction before the adverbial, which we excised and replaced with a gap. The other half lacked a conjunction before the adverbial (*implicit passages*). With these, we simply inserted a gap before the adverbial, so that all passages had the following structure (also see Figure 2a):

(5) *Bruce, who was in Edinburgh at the time, was in the audience on the opening night _____ afterwards the Director invited Bruce to join him and some members of the cast for a drink in a pub in the Grassmarket.*

For each of the 20 adverbials used in the study, participants saw 25 *explicit passages* and 25 *implicit passages*, with the exception of *however*, which rarely occurs immediately after a conjunction. For

however, we were only able to include 25 implicit passages and 1 explicit passage in the study. (Note that passages from one or more of our studies are given in italics. Examples which simply illustrate a point are presented in standard font.)

As for results, responses on *explicit passages* showed that participants selected the authors' original conjunctions 57% of the time. If we take participant BUT as substitutable for author AND in the context of the passage (and likewise, participant SO and author AND), agreement increases to 70%. Divergences with authors' original conjunctions provide evidence of certain adverbials having a preference for certain conjunctions, although it is neither the case that all adverbials co-occur with the same preferred conjunction, nor the case that each adverbial has a single preferred conjunction. However, responses to the *implicit passages* also demonstrate patterns that are unique to certain adverbials.

As with *explicit passages*, no single conjunction is preferred across the board with *implicit passages*, nor a single conjunction preferred uniformly for a given adverbial. Despite this non-determinism, pockets of systematicity arise. In some cases, similar adverbials show similar preferences: e.g., the pairs *nevertheless/nonetheless* and *therefore/thus* show a preference for the conjunction BUT and SO respectively. The variability that emerges is often passage-specific: e.g., some passages with *instead* favored BUT, while others lent themselves to the inference of BECAUSE. Further information and discussion can be found in (Rohde et al., 2016).

The second study, described in (Rohde et al., 2017), covered 37 discourse adverbials and used participants (N=28) recruited locally rather than via Amazon Mechanical Turk (AMT), so as to better ensure participants would complete the study. To ensure there was no dramatic difference between the AMT participants in the first study and local participants in the second, we carried out a pre-trial test on thirty-six passages from the first study: 18 passages where most AMT participants chose the same response (*strong signal*) and 18 passages on which their' responses spanned more than one response (*divided signal*).

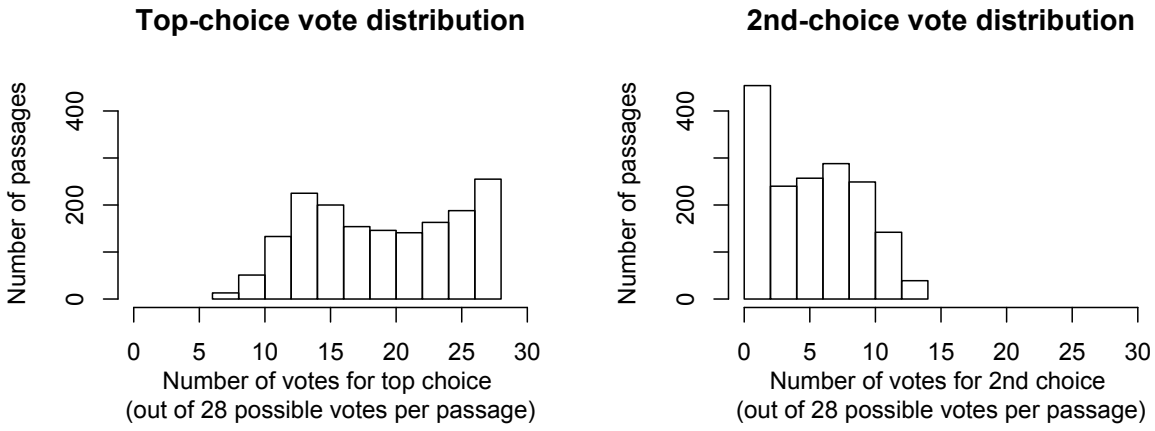
We found that local participants agreed with AMT participants on the *strong signal* passages; on the *divided signal* passages, responses of local participants also varied, although not necessarily showing the exact same response pattern. A comparison across all 36 passages showed the response profiles of all items to be highly correlated across the two participant groups. So we concluded that the results of the two studies would be comparable, even with the new participant pool.

In addition, to reduce the prevalence of ambiguous and/or less informative responses (that is, AND and the no-conjunction response NONE), we offered these two possibilities as options only if a participant chose the response OTHER. This did indeed reduce the frequency with which AND and NONE were used.

As for results, for explicit passages, agreement with the author's original choice is comparable to the first study at 53%, rising to ≈70% if AND is taken to be *conditionally substitutable* for BUT and for SO. For the implicit passages, we see striking differences between adverbials. Some favor a conjunction that conveys a similar sense: e.g., *consequently*, *as a result*, *accordingly*, and *hence* all favor SO. Other preferences reflect usage rather than semantics. For example, *for one thing*, *first*, and *after all*, together with their context of use, favor the conjunction BECAUSE. From this, it is possible to see that adverbials must be characterized in terms of both their own semantics and their use in context.

2.2.2 Competition between (substitutable?) connectives

Of particular relevance to the current study (Section 3) is that participants showed strong biases about which conjunction(s) they saw as best expressing their interpretation. This can be seen in how frequent the top choice and second choice is for each passage in the second study: Figures 1(a) and 1(b) show that the top choice is typically favored by more than half the participants, whereas the second choice is favored by fewer than half. If all choices were near equi-probable, the top choice would have only achieved a plurality rather than a majority. If the selections that differed from the favored one were just noise, we would have expected to see the histogram pushed much farther to the left in Figure 1(b). However, the second choice frequently receives 5-10 votes, suggesting that when a passage permits multiple conjunctions to be selected across participants, there is consistency in those additional selections. This raises the



(a) (b)

Figure 1: (a) Across passages, how many participants (out of 28) favored top choice? Histogram of the number of votes received by the favored conjunction in each passage. (b) Across passages, how many participants favored second choice? Histogram of number of votes given to second conjunction choice

question of what patterns underlie the pairings of the top-choice and runner-up connectives for a given passage and whether those pairings reflect distinct interpretations of the passage or the substitutability of those connectives in that context.

3 Substitutability Study

An important question raised by the earlier studies in this project is whether differences in participant choices indicate differences in understanding (*alternative interpretation*) or simply differences in how best to express otherwise shared understanding (*alternative preferences*). To answer this question, we devised a task to explicitly elicit participant judgments on all candidate connectives that could express their understanding of the passage.

3.1 Participants

Participation in the task was limited to participants from the second study (Section 2.2), who would thereby be familiar with the connective insertion task. Of the original 28 participants, we were able to collect data from 16 (11 female; ages 19-69 (mean 35); highest education: 5 high school, 6 undergraduate degree, 3 masters, 2 PhD).

3.2 Materials

We selected 67 passages for the task from passages used earlier; predominantly, but not exclusively, passages used in the first (Amazon Mechanical Turk) study. Both explicit and implicit passages were included, based on whether earlier responses to a passage would help explore whether divergent connective preferences reflected divergent readings or substitutability. For example, Ex. 6 (an implicit passage, with no connective adjacent to *therefore*) had earlier received responses split between AND and SO, whereas Ex. 7 (an explicit passage with author BECAUSE) had received responses split between BECAUSE and BUT.

(6) *Neocons pushed for this war _____ therefore they deserve the blame for its failure or the credit for its success.*

(7) *"Nervous? No, my leg's not shaking," said Griffey, who caused everyone to laugh _____ indeed his right foot was shaking.*

We speculated that the response pattern to passages like Ex. 6 might simply reflect participants choosing their preferred conjunction from two essentially substitutable options (AND and SO), while the response pattern to passages like Ex. 7 might reflect different interpretations, with BECAUSE linking the final segment to "caused everyone to laugh", and BUT contrasting it with "... my leg's not shaking". This would be in line with Knott's predictions: AND and SO are contingent substitutable, in which case participants asked to select conjunctions with the same meaning might select both AND and SO for Ex 6. On the other hand, with BUT and BECAUSE, participants would be unlikely to select both as expressing the same meaning.

Of the 67 passages, 46 were selected to test two of Knott's contingent substitutability claims (see breakdown in Figure 2b). The substitutability of AND and SO was tested with passages containing the discourse adverbials *for example, therefore, afterwards* or *then*, while the substitutability of AND and BUT was tested with passages containing *in fact, in general, (more) specifically* or *meanwhile*.

We also targeted connective combinations which are not substitutable under Knott's analysis and hence would be predicted to yield *exclusive* response patterns. These included 6 passages that had previously shown a combination of BECAUSE and BUT responses (with the discourse adverbials *after all*, *previously*, and *indeed*); 3 that had previously shown a combination of BECAUSE and OR responses (with the adverbials *otherwise* (2) and *hence* (1)); and 2 that had previously shown a combination of SO and OR responses (with adverbial *in other words*). These were to be contrasted with one passage that had previously received almost uniform BECAUSE responses (with adverbial *in fact*).

A further 8 passages targeted the insertion of no connective. These passages had received frequent NONE responses in our previous work and were included here so that not all passages would necessarily require the insertion of any/many connectives. Finally, we also included two "catch trials" in the form of constructed examples for which there was logically only one answer (e.g. OR for *David weighs more than Alice _____ Alice weighs more than David*).

3.3 Procedure

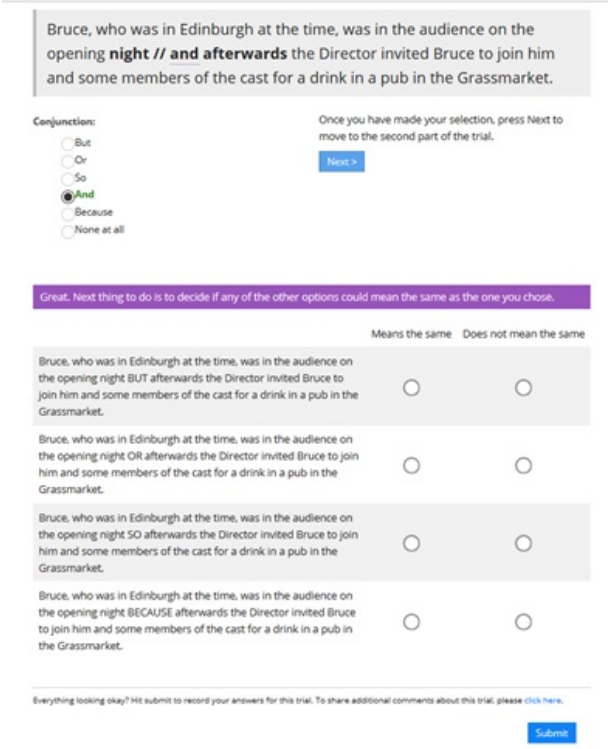
Participants were shown two text spans and asked to select the conjunction that best described the relationship between the spans. There were five available conjunctions: AND, BECAUSE, BUT, OR, and SO. *None at all* was also an option. After selecting their best choice, participants were shown a list of sentences with the gap replaced by each of the remaining conjunctions, and told: "Next thing to do is to decide if any of the other options could mean the same as the one you chose." Next to each sentence were two radio buttons: "Means the same", and "Does not mean the same". Participants had to select one of the buttons for each sentence before they could proceed to the next question. The screenshot in Figure 2a shows the layout of the interface. The order of candidate connectives and of passages was pseudo-randomized to control for order effects.

4 Results

For each passage, we assessed each of the 16 participants' response profiles and categorized them as responding with only one connective (e.g., "only SO") or a best plus other choice(s) (e.g., "SO:AND") or a best plus other choices that did *not* include some other relevant connective (e.g., "SO:¬AND"). This evaluation allows us to see which passages were dominated by which response profiles.

For example, in (Ex. 6), only one participant rejected the full substitutability of AND/SO. The 16 participants' response profiles consisted of 11 "SO:AND", 4 "AND:SO", and 1 "only SO". In (Ex. 7), as predicted, BECAUSE/BUT substitution was rare. The response profiles consisted of 11 "only BECAUSE", 1 "only BUT", 3 "BECAUSE:AND", and 1 "BECAUSE:AND,BUT".

(N.B. Our two catch-trials showed only one participant selecting an incorrect response. Since this participant's responses on target trials were in line with the other participants, their data was not excluded.)



(a)

AND/BUT substitutability (n=22)

prior behavior	#
mostly AND	6
mostly SO	8
AND/SO mix	8

AND/BUT substitutability (n=24)

prior behavior	#
mostly AND	8
mostly SO	8
AND/SO mix	8

(b)

Figure 2: (a) Interface: first conjunction selection indicates participant's first choice; subsequent versions of the passage are then displayed with alternative conjunctions. (b) Materials: Distribution of passages according to their behavior in our prior studies.

4.1 Predictions of contingent substitutability

Recall that evidence for *contingent substitutability* between two connectives involves both (1) response profiles in which a participant licenses either one of the connectives but not the other and (2) profiles in which a participant licenses both. To examine whether participants' behavior reflects these two scenarios, we consider the relationship between AND/BUT and AND/SO.

For the analysis, we visualize the distribution of participant response profiles for the passages relevant to the substitutability of AND/BUT (n=24) and AND/SO (n=22). In Figures 3 and 4, each bar represents the response profiles for a single passage. Bar length indicates the number of participant response profiles (out of 16) which adhere to the contingent substitutability hypothesis. The colorful part of the bar (above the x-axis) shows response profiles satisfying the free-exchange aspect of contingent substitutability and the one-but-not-both aspect of contingent substitutability. The grey part of the bar (below the axis) shows response profiles in which the participant selected as their first choice one of the relevant connectives but then selected an unexpected connective as their second choice. Short bars correspond to passages whose response profiles are not relevant to the AND/BUT contingent substitutability hypothesis (i.e., where the first choice of many of the participants was neither AND nor BUT).

4.1.1 Substitutability between AND/BUT

Because the connectives AND and BUT are related in Knott's taxonomy via *contingent substitutability*, we check the response profiles of passages that previously had shown a mix of AND/BUT responses across participants. Figure 3 shows the distribution of response profiles. Some passages indeed favored BUT as the first or only choice (red bars on the left); others favored AND as the first or only choice (blue bars on the right). Crucially, as Knott predicts, most passages yielded response profiles fulfilling the two realizations of *contingent substitutability* (colorful tall bars). For a minority of passages, participants

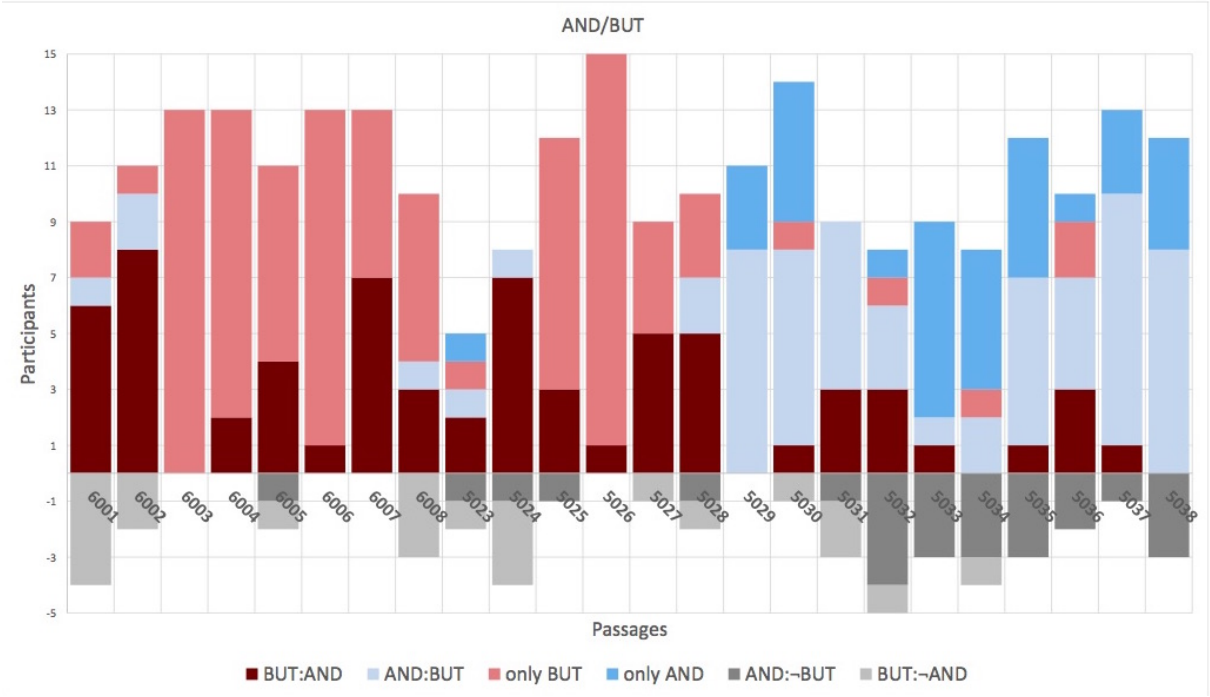


Figure 3: Distribution of response profiles for passages testing AND/BUT substitutability. Positive portion of bars (above the x-axis) corresponds to response profiles in keeping with contingent-substitutability predictions; negative portion (below the x-axis) corresponds to response profiles with unexpected substitutions.

diverge from the predicted patterns by selecting non-AND/BUT options like BECAUSE/SO/OR. We return in Section 4.2 to more systematic cases that violate this type of exclusivity.

4.1.2 Substitutability between AND/SO

Knott's taxonomy also relates AND/SO via *contingent substitutability*. We check the response profiles of passages that previously had shown a mix of AND/SO responses. Figure 4 shows the distribution of response profiles. Some passages favored SO as the first or only choice (dark blue/yellow bars on the left); others favored AND as the first or only choice (light blue/turquoise bars on the right). Crucially, as Knott predicts, most passages yielded response profiles fulfilling both realizations of contingent substitutability (colorful tall bars).
Figure 4 shows that there are some cases (e.g., the large gray negative bar for passage 5019) where the hypothesis that AND alternates with SO was not upheld. In that particular passage (Ex. 8), the adverbial *then* yielded an alternation between AND/BUT rather than AND/SO (akin to the pattern in Figure 3).

(8) *A bone-marrow transplant is a medical resurrection. First doctors all but kill a patient _____ then they bring him back to life.*

This finding is not that surprising since we selected passages to test the AND/SO hypothesis that contained adverbials that had previously permitted a mix of AND and SO, but (Ex. 8) had previously yielded a dominant AND bias and did not lend itself to a RESULT inference. This confirms the observation from (Rohde et al., 2017) that even though some adverbials introduce preferences regarding the implicit relation participants infer, the content of the passage is crucial as well.

4.2 Predictions of exclusivity for connectives

Knott also predicted that particular pairs of connectives are *exclusive*, meaning that they cannot substitute for one another in any context and convey the same sense (Section 2.1). Our current study provides

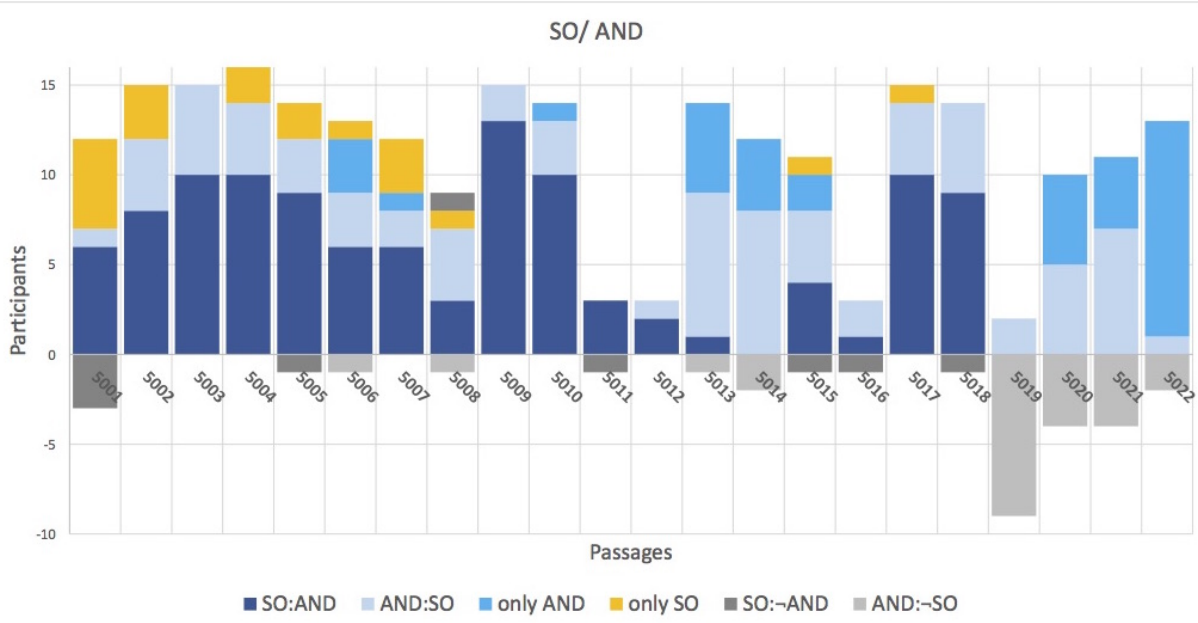


Figure 4: Distribution of response profiles for passages relevant to the hypothesis of AND/SO substitutability. Figure properties (above and below the x-axis) match those of Figure 3.

evidence against such exclusivity. While taken as a whole, the results of our study may not surprise the reader, they allow one to start pursuing a better explanation of what discourse connectives do and how they do it, and how one may need to alter one's predictions in the presence of a discourse adverbial.

Here we examine pairs of connectives that should be *exclusive*: In some cases, members of the pair belong to only one sense category, but the categories differ. In other cases, at least one member of the pair appears in multiple sense categories but nonetheless fails to share a category with the other. We discuss possible explanations for two of these in the Discussion in section 5.

because/but BECAUSE only belongs to the category CAUSE, while BUT belongs to the sense category NEGATIVE POLARITY. As such, they should not be *substitutable*. Nevertheless, the response profiles for Ex. 9–10 (involving the adverbials *previously* and *after all*) showed that at least half of our participants endorsed both BECAUSE and BUT, with a slight preference for BECAUSE as first choice and BUT as the other connective expressing the same meaning. The remaining responses consisted of "only BECAUSE" and "only BUT" profiles (plus two "BUT:AND" profiles, a pairing already highlighted as contingently substitutable and confirmed in our results in Section 4.1.1). This pattern of response profiles suggests that BECAUSE/BUT are contingently substitutable in the context of *previously* and *after all*.

(9) *The demand for tickets continued so strong yesterday that several carriers, including United and Delta, extended until Sunday the period in which customers must pick up the tickets they had booked by telephone. _____ previously they were required to do so within 24 hours.*

(10) *Yes, I suppose there's a certain element of danger in it, that you can't get around _____ after all, there's a certain amount of danger in living, whatever you do.*

because/so BECAUSE only belongs to the category CAUSE, while SO belongs to the sense category RESULT. As such, they should not be *substitutable*. Nevertheless, the response profiles for Ex. 11–12 involving the adverbial *then* showed that nearly all participants endorsed both BECAUSE and SO (with a strong preference for BECAUSE as first choice and SO as the other connective expressing the same meaning). This was the case for all 16 participants for Ex. 11 and 14 out of 16 for Ex. 12, with a 15th participant who selected AND as their first choice in an "AND:BECAUSE,SO" profile. This pattern of response profiles suggests that BECAUSE/SO are not always exclusive.

(11) *With a $50 credit in an on-line account, Jordan eagerly logged on. But as he tried to decide which video games to buy, he realized he had a new problem: shipping costs put him over budget. It took him a few weeks to figure out a solution: when he finally made his first purchase in July, he opted for less expensive items – videotapes – _____ then he could afford to pay the shipping costs.*

(12) *On a sunny day, upward of 2,000 hot dog vendors are at large in New York. Like many veteran frankfurter men, Mr. Stathopoulos is his own boss. There are a number of small companies who have a dozen or so carts and hire people to operate them, paying them a salary. All hot dog men, though, dream of owning a cart _____ then they can set their own hours and the harder they work the more money they make.*

but/or BUT belongs to one category, NEGATIVE POLARITY, while OR belongs to multiple categories: SEQUENCE (where it is a synonym for 'or else'), RESTATEMENT (where it is a synonym for 'or rather'), and NEGATIVE POLARITY (no synonyms). As such, BUT/OR should only be *substitutable* in contexts in which a *negative polarity* relation is operative. However, in the context of the discourse adverbial *more specifically*, BUT/OR/AND are substitutable as examples of RESTATEMENT. The response profiles for Ex. 13–14 showed that BUT/OR is endorsed by 10 and 16 of the 16 participants, respectively. A number of participants also add AND, which is contingently substitutable for BUT.

(13) *Windows is a way of life to some degree _____ more specifically it's Microsoft's way of life, and you'd better like to live the way they tell you to live, or else.*

(14) *"The Wild Hawaiian" is a Hawaiian rock album _____ more specifically it's an album of songs in the Hawaiian language, against a whiplash of percussion and distorted guitars.*

or/because As noted, BECAUSE belongs to only the sense category CAUSE, and does not overlap with any of the categories that OR belongs to. Passages Ex. 15–16 contain *otherwise*. The response profiles showed that 14 of 16 participants endorse OR/BECAUSE for Ex. 15 while 12 of 16 do so for Ex. 16.

(15) *"If people want to get rid of an animal, they'll do so," she said. "My thing is to get the animal here _____ otherwise they're going to end up wandering in the streets."*

(16) *Gouges are deep scratches that must be filled as well as colored _____ otherwise they will collect dirt and become permanently discolored.*

or/so Like BUT and BECAUSE, SO belongs to only one sense category (RESULT), and does not overlap with any of the OR categories. Passages Ex. 17–18 contain *in other words*. The response profiles showed that 10 of 16 participants endorse OR/SO for Ex. 17, and 13 of 16 do so for Ex. 18.

(17) *In recognizing their special responsibilities and working sensitively in developing countries, multinationals can expect a smoother and more sustained market development in the long run _____ in other words good ethics is good business.*

(18) *Unfortunately, nearly 75,000 acres of tropical forest are converted or deforested every day _____ in other words an area the size of Central Park disappears every 16 minutes.*

We stress that the above results depend crucially on having sufficient participants to show more than one strong pattern of responses. With only a few participants, this might simply look like noise.

5 Discussion

The results we report here demonstrate a wider range of substitutability than previously assumed, but the experiments themselves do not answer why such a range is possible and what its limits are likely to be. In speculating about possible explanations of our observed patterns of substitutability, the picture is complicated – we have not found a uniform explanation that holds across the board to account for all passages for all adverbials. Rather, the substitutability patterns of families of different adverbials appear to

invite different explanations, where those families of adverbials reflect shared patterns of substitutability that can cross-cut their semantics and function.

Below we review two patterns we see with regards to conjunction alternations, where the minimal explanation for the observed patterns of alternations appears to involve *both* the coherence relation signalled by the discourse adverbial *and* an additional coherence relation derived through pragmatic inference.

5.1 Disjunction and alternatives

Both the adverbials *otherwise* and *in other words* were found to license substitutability of the conjunction OR – specifically OR/BECAUSE in the presence of *otherwise* and OR/SO in the presence of *in other words*. We speculate that this substitutability arises due to the fact that both adverbials encode 'otherness' in their lexical semantics, as well as in their surface forms. Because these adverbials always convey an alternative, this aspect of their meaning can be made explicit with OR (albeit redundantly) whenever they appear. Then if an *otherwise* passage also supports inference of causality, OR can alternate with BECAUSE. The passage in Ex. 15 allows such a causal inference: The reason for the speaker's particular goal is BECAUSE the alternative is worse.

However, *otherwise* passages that do not support a causal inference do not permit this alternation. For example, the constructed *otherwise* passage in Ex. 19 lists a set of alternatives without describing a reason for a particular course of action; in that case, no OR/BECAUSE alternation is predicted to arise.

(19) For dinner, sometimes we go to a restaurant or to visit friends _____ otherwise we eat at home.

A related pattern emerges for *in other words*. The sense of an alternative is implicit in the adverbial itself since its meaning is about reformulation, thereby licensing OR in all cases. But OR can be seen to alternate with SO in the context of *in other words*, as shown in Ex. 20: The generalization about the rate of deforestation has the consequence that a more specific reformulation of that calculation must also be true.

(20) *Unfortunately, nearly 75,000 acres of tropical forest are converted or deforested every day _____ in other words an area the size of Central Park disappears every 16 minutes.*

We speculate that this OR/SO alternation for *in other words* arises because of the nature of reformulation: When one formulation holds, it follows that the alternative formulation must also hold. In this way, *otherwise* and *in other words* are similar in permitting the alternation between OR and a causal conjunction. They differ in that *otherwise* allows only conditional substitutability between OR/BECAUSE, whereas *in other words* may always license the alternation.

5.2 Apparent Symmetry in Causality

We normally assume that causality is asymmetric: If X BECAUSE Y (i.e., reason), then it can't also be the case that X SO Y (i.e., result). Nevertheless, as we noted in Section 4.2, the response profiles for Ex. 11–12 involving the adverbial *then* showed that nearly all participants endorsed both BECAUSE and SO for the same passage.

In this case, the explanation revolves around both an inference of purpose and a fact that we hadn't considered – that SO can be ambiguous between conveying result and conveying purpose. For example, in the constructed example Ex. 21a, SO can convey purpose, while in the minimally different Ex. 21b, it conveys only result.

(21) a. Nathan renewed his passport this year so (that) he could travel abroad.

b. Nathan lost his passport this year so (therefore) he could not travel abroad.

While we were aware that the purpose sense of SO can be expressed by the explicit phrase SO THAT, we had ignored the fact that this can (and is) often reduced to simply SO.

This sense of purpose can be inferred in both Ex. 11–12. However, it is not associated with the adverbial *then*, but rather the modal (*could* in Ex. 11, and *can* in Ex. 12). Thus we would predict that, whatever the discourse adverbial, if purpose can be inferred, then participants will identify BECAUSE and SO as substitutable.

There are other circumstances in which SO/BECAUSE can alternate – in particular, the epistemic uses of these conjunctions that capture reasoning about WHY a conclusion can be drawn or what conclusion can be drawn as a RESULT of certain evidence. The alternation depends on reasoning about the events that happened in the world versus reasoning about how a speaker's conclusion is drawn. Consider the well-known example Ex. 22 from Van Dijk (1977) (see also earlier work by Rutherford (1970)):

(22) a. John is home so/because the lights are burning.

b. The lights are burning so/because John is home.

The conjunction SO primarily conveys consequence, whereas BECAUSE primarily conveys cause. So if the order is "Real-world-cause ____ Real-world-effect", as in Ex. 22a, then SO will express real-world causality, while BECAUSE will express epistemic inference (the opposite causal direction: reasoning backwards). Conversely, if the order is "Real-world-effect ____ Real-world-cause", as in Ex. 22b, then BECAUSE will express real-world causality, while SO expresses epistemic inference.

6 Conclusions and Future Plans

The experiment reported here collected multiple judgments from multiple participants on naturally-occurring passages containing discourse adverbials. The data provide evidence confirming two of Knott's *contingent substitutability* pairs and denying several *exclusivity* pairs. We have offered reasons for three of the surprising non-exclusivity judgments: for OR/SO, OR/BECAUSE, and BECAUSE/SO. But there is clearly more to explain and more evidence to gather in support for possible explanations. The question of how different discourse connectives are used to realize particular types of coherence relations remains unresolved. We plan to continue our pursuit of an answer via experimentation and classification of patterns of use in these contexts.

Acknowledgments

Our thanks to Alexander Johnson for enlightening in-person discussion, to Ali Knott for enlightening discussion over email, and to Larry Horn for pointing us to the work of Rutherford (1970). This project has been supported in part by a grant from the Nuance Foundation.

References

Knott, A. (1996). *A Data-driven Methodology for Motivating a Set of Coherence Relations*. Ph.D. dissertation, Department of Artificial Intelligence, University of Edinburgh.

Knott, A. and C. Mellish (1996). A feature-based account of the relations signalled by sentence and clause connectives. *Language and Speech 39(2-3)*, 143–183.

Rohde, H., A. Dickinson, C. Clark, A. Louis, and B. Webber (2015). Recovering discourse relations: Varying influence of discourse adverbials. In *Proceedings, First Workshop on Linking Computational Models of Lexical, Sentential and Discourse-level Semantics*, Lisbon, Portugal, pp. 22–31.

Rohde, H., A. Dickinson, N. Schneider, C. Clark, A. Louis, and B. Webber (2016). Filling in the blanks in understanding discourse adverbials: Consistency, conflict, and context-dependence in a crowdsourced elicitation task. In *Proceedings of the Tenth Linguistic Annotation Workshop (LAW-X)*, Berlin, pp. 49–58.

Rohde, H., A. Dickinson, N. Schneider, A. Louis, and B. Webber (2017). ConnText: Recognizing concurrent discourse relations. Second annual report to the Nuance Foundation.

Rutherford, W. E. (1970). Some observations concerning subordinate clauses in English. *Language 46*, 97–115.

Sandhaus, E. (2008). New York Times corpus: Corpus overview. LDC catalogue entry LDC2008T19.

Spooren, W. (1997). The processing of underspecified discourse relations. *Discourse Processes 24*, 149–168.

Van Dijk, T. A. (1977). *Text and Context*. London: Longmans.